BJPsych | The British Journal of Psychiatry (2018) 212, 301–307. doi: 10.1192/bjp.2018.9

Sensitivity and specificity of the Akena Visual Depression Inventory (AViDI-18) in Kampala (Uganda) and Cape Town (South Africa)

Dickens Akena, John Joska and Dan J. Stein

Background

Visual scales may be particularly useful in screening for depression in patients with low literacy. However, few have been validated and none are in common use.

Aim

Modification and validation of a visual scale to screen for depression in low literacy settings.

Method

We assessed the validity, reliability and factor loading of a 28-item visual depression inventory using pictorial items depicting depression signs and symptoms. We validated a revised scale comprised of 18 items known as the Akena Visual Depression Inventory (AViDI-18) against a structured diagnostic interview (Mini-International Neuropsychiatric Inventory) in 343 patients in Kampala (Uganda) and Cape Town (South Africa).

Results

The 18 pictorial items had acceptable validity and reliability. The area under the curve (AUC) score of the AViDI-18 was 0.9. AUC scores were not significantly associated with sociodemographic variables.

Conclusion

The AViDI-18 is a valid screen for depression in patients with low literacy.

Declaration of interest

None.

Depression is highly prevalent and burdensome, with a particularly wide treatment gap apparent in low- and middle-income countries.[1,2] A number of clinician- or self-rated depression symptom severity measures, including the Hamilton Depression Rating Scale,[1] the Centre for Epidemiologic Studies-Depression[2] and the Patient Health Questionnaire (PHQ-9),[3] have been used to screen for depression. The administration of clinician-rated symptom severity measures is time consuming, whereas many self-rated instruments can only be administered in high literacy contexts. Given that up to 33% of people in sub-Saharan Africa are unable to read or write,[4] it has been suggested that visual scales may be particularly useful in this setting.[5,6]

Research examining the accuracy of visual scales in screening for depression has been evolving since the late 1960s. The majority of such scales were developed for use in patients with cognitive deficits.[7–10] Results from these studies have been conflicting, with some recommending the use of visual scales[5,6,9] and others not.[7,10] One possibility is that scales with fewer items have lower specificity;[11,12] after all, depression is a condition characterised by multiple signs and symptoms.

Preliminary work

In earlier work, we developed a 14-item visual scale consisting of items depicting signs and symptoms of depression, the Akena Visual Depression Inventory (AViDI),[5] and validated it against the Mini-International Neuropsychiatric Inventory (MINI).[13] This work had a number of limitations, including no examination of the validity (face, content and criterion) and reliability (internal, inter-rater and test-retest) of the included items. This study addresses these limitations by (a) examining the validity (construct, content and face) and reliability (internal, test-retest and inter-rater) of visual scale items, and (b) assessing the sensitivity and specificity of the modified AViDI against the MINI in Kampala (Uganda) and Cape Town (South Africa). The two populations studied here differ in literacy level, socioeconomic status and ancestral origins.[14] We therefore also examined whether accuracy of the modified AViDI was affected by sociodemographic variables.

Method

Study sites

The study sites for the investigation of validity and reliability of the items for inclusion in the visual scale were the Mityana Hospital and the Mildmay clinic in Uganda, and the out-patient clinic of the Department of Psychiatry at Groote Schuur Hospital in South Africa. Mityana Hospital is 65 km west of Kampala city and serves mainly rural and peri-urban populations. The Mildmay clinic is 10 km south of Kampala and mainly serves an urban population. Groote Schuur Hospital is 5 km west of Cape Town and serves mainly an urban population.

The study sites for the work on sensitivity and specificity (validation) of the visual scale were the Mityana Hospital and Butabika National Referral Mental Hospital in Uganda, and the MZM clinic in Philippi, Cape Town. Butabika Hospital is 10 km east of Kampala and serves all populations (rural, peri-urban and urban populations). The MZM clinic is 20 km south east of Cape Town and serves mainly a peri-urban population. All patients provided informed consent before participating in the study.

Sample size and power calculations

The sample size for the investigation of validity and reliability was based on previous validity and reliability studies, where sample sizes ranged from 300 to 500 participants.[15,16] Burderer's formula[17] was used to calculate the sample size for the specificity and sensitivity study. With an estimated 12-month prevalence of 30% among

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press

attendees of out-patient clinics in sub-Saharan Africa,[18] 353 participants – of which 105 would have depression – are needed to achieve a sensitivity of 80%, providing 80% power with a 95% confidence interval.

Objective 1: validity and reliability assessment

Assessment of the validity and reliability of the pictorial items of the visual scale comprised the following six steps:

Step i: redesigning representation of items for inclusion in the scale

A fine artist at Makerere University provided illustrations depicting the following eight DSM-based depression signs and symptoms: sadness, anhedonia, low appetite, insomnia, low energy, crying spells, extreme worry and suicidality. In our previous work, we had used the same signs and symptoms, with the exception of insomnia.[5]

Step ii: construct and content validity assessment

We conducted in-depth qualitative interviews to assess construct and content validity by showing the 28 pictures (drawn in step i) to 16 psychiatrists and clinical psychologists with at least 10 years of experience in practicing general adult psychiatry in Uganda and South Africa. These clinicians were asked to indicate if the pictures were an accurate representation of the relevant signs and symptoms, as found in their patients. The Lawshe method for assessing content validity was used by asking the participants to describe whether the pictures were: (a) essential, (b) useful but not essential, or (c) neither useful nor essential in assessing depression.[19] The interviews were tape recorded, transcribed and entered into NVivo version 16 (run on Windows 7) for exploration of key themes and relationships between themes.

Step iii: face validity

To assess whether the items were a representation of what they were intended to measure (face validity), we presented 24 pictures (generated upon completion of step ii) to 10 individuals who had been diagnosed with a severe major depressive disorder (MDD) at the Butabika Hospital and who were in clinical remission and due for discharge. Participants were asked to state whether the pictures depicted how they felt while suffering from the MDD, to suggest modifications where there was lack of clarity, or to suggest whether new pictures were required to depict their depression signs and symptoms. We chose individuals with depression to assess for face validity as persons who have never suffered from a depressive disorder would be unlikely to identify with the depicted items; such individuals would simply state that they have never felt that way, information that would not be very useful.

Step iv: refining scale items

The results from the qualitative interviews in steps i–iii were used to further refine and improve the pictures. The fine artist was contacted to redraw some of the items based on answers to the questions above.

Step v: administration of the pool items to a developmental sample

Trained bachelor's degree-level research assistants then administered the 24 pictorial items to a random sample of 333 participants at Mityana Hospital out-patients department (OPD), the Mildmay clinic and Groote Schuur Hospital. Eligibility criteria included participants who were at least 18 years old with no overt psychiatric or physical illness that would require urgent attention (including hospital admission) as clinically assessed by nurses at triage. Each of the pictures was presented to participants, who were then asked whether the particular picture depicted how they felt over a 2-week period. If the respondent endorsed no, they would get a score of zero. Participants who answered yes would then be asked to state whether, over the past 2 weeks, they felt that way sometimes (scored 1), most of the time (scored 2) or almost all the time (scored 3). Participants were asked to indicate whether there was lack of clarity of the items, and if so to suggest how to improve the items. Sociodemographic data were collected.

Step vi: assessing reliability

We conducted test-retest reliability by administering the items to participants ($n = 30$) at baseline and then to the same participants 1 week later. To test the inter-rater reliability, two research assistants administered the items to participants ($n = 30$) an hour apart. The two research assistants occupied different rooms and were blinded to the each other's results.

Data analysis

STATA version 12.1 (run on Windows 7) was used in the analysis of the quantitative data obtained from steps v and vi. We used linear regression analyses to examine for associations between item scores and sociodemographic variables (age, gender, education level, employment and marital status) and HIV status. Factors that were statistically significant at bivariate analysis were then entered into a hierarchical multivariable linear regression model. We assessed for inter-item correlations, scale reliability coefficient (Cronbach's α), test-retest reliability, inter-rater reliability and factor loading analyses. We used the principal components method at factor loading analysis, retaining factors with an eigenvalue >1.2; eigenvalues >1.2 are indicative of more than a single construct in a scale.[20]

Objective 2: validating (assessing the sensitivity and specificity) the AViDI-18

On the basis of the results from steps i–vi, 18 items were included in the modified AViDI-18 for comparison against the MINI gold standard.

This part of the study was conducted at the Mityana Hospital OPD ($n = 100$), at Butabika Hospital OPD ($n = 107$) and at the MZM clinic in Cape Town ($n = 136$). All participants were interviewed using the AViDI-18 and the MINI by two independent research assistants who were blinded to each other's results. Half of the participants had the AViDI-18 administered first, and the other half had the MINI administered first. Participants diagnosed with MDD according to the MINI were referred to the mental health clinic for treatment by the mental healthcare practitioner.

Using STATA version 12.1, we conducted factor loading analyses of the final items and calculated the sensitivity, specificity, likelihood ratios and predictive values of the AViDI-18 in relation to the MINI.[19] We also computed the area under the receiver operating characteristic curve (AUC) score, which is a summary statistic (measure of scale accuracy) that is a function of the sensitivity, specificity, likelihood ratios and predictive values. AUC scores were compared across sociodemographic variables (age, gender and level of literacy), HIV status and country.

Results

Objective 1: validity and reliability assessment of the scale items

Steps i–iv (face, construct and content validity)

Of the 28 items, 4 were judged irrelevant on the basis of the Lawshe Method (described in step ii above) and were removed. The

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press

Table 1 Baseline demographics of participant characteristics for objective 1

Study variable	Sociodemographic variables	*n* (%)	Bivariate analysis β-coefficient (95% CI) (item scores)	Multivariable analysis β-coefficient (95% CI) (item scores)
Gender	Male	84 (25.23)	−0.09 (−3.91–1.92)	N/A
	Female	249 (74.77)	*P* = 0.502	
HIV status	Positive	189 (56.76)	0.41 (−1.25–2.08)	N/A
	Negative	144 (43.24)	*P* = 0.623	
Age	18–30	214 (64.26)	0.23 (0.12–0.34)	0.17 (0.05–0.29)
	31–49	99 (29.73)	*P* < 0.001	*P* = 0.03
	>50	20 (6.01)		
Education	No formal education	19 (5.71)	−4.29 (−5.87 to −2.71)	−3.11 (−4.77 to −1.44)
	Primary level education	116 (34.83)	*P* < 0.001	*P* < 0.001
	Secondary level education	115 (46.55)		
	Post-secondary level education	43 (12.91)		
Marital status	Married	147 (44.14)	1.79 (0.61–2.96)	0.95 (−0.22–2.12)
	Never married	117 (35.14)	*P* = 0.003	*P* = 0.113
	Widowed	15 (4.50)		
	Separated/divorced	54 (16.22)		
Employment	Formal employment	117 (35.14)	0.60 (0.06–1.13)	0.65 (0.13–1.17)
	Informal employment	106 (31.83)	*P* = 0.027	*P* = 0.013
	Unemployed	110 (33.03)		
Study site	Mildmay Lweza	167 (50.15)	1.18 (0.58–3.11)	0.92 (−0.62–3.16)
	Mityana	116 (34.83)	*P* = 0.033	*P* = 0.522
	Butabika	50 (15.01)		

N/A, multivariable analysis not applicable since variables were not significant at bivariate analysis.

removed items included two pictures depicting happiness and two pictures depicting fear, which participants found ambiguous. The clinicians also suggested a number of modifications. Participants who had suffered from MDD also made suggestions about improvements, including 'enhancing' smiles on the faces, making the teeth/eyes more visible, using 'neutral' colours (not extremely bright or dark) for clothing and hanging by ropes as the most relevant method of dying by suicide (rather than guns or overdosing on medicines).

Step v (construct validity and reliability assessment)

The 24 items were administered to 333 participants, 84 of which were male (25.2%). The mean age of participants was 29.7 years (s.d. ± 11.4), approximately two-thirds (214, 64.3%) of whom were below 30 years old. Over one-third (135, 40.6%) had low levels of education (no formal education or only up to primary level education) and approximately half (147, 44.1%) were married. About one-third (119, 35.7%) of participants were formally employed. More than half (189, 56.7%) of the participants were HIV-positive.

Table 2 Item correlations for objective 1

Item	Correlation	Item test correlation	Item rest correlation	Average inter-item correlation	α
1	Positive	0.6794	0.6385	0.3143	0.9134
2	Positive	0.7539	0.7208	0.3101	0.9118
3	Positive	0.7004	0.6616	0.3131	0.9129
4	Positive	0.7113	0.6736	0.3125	0.9127
5	Positive	0.7602	0.7278	0.3098	0.9117
6	Positive	0.7081	0.6700	0.3127	0.9128
7	Positive	0.5752	0.5250	0.3201	0.9155
8	Negative	0.3155	0.2498	0.3347	0.9204
9	Positive	0.5697	0.5190	0.3204	0.9156
10	Positive	0.5918	0.5429	0.3192	0.9151
11	Positive	0.7370	0.7021	0.3111	0.9122
12	Positive	0.6968	0.6576	0.3133	0.9130
13	Positive	0.6237	0.5776	0.3174	0.9145
14	Positive	0.7117	0.6741	0.3125	0.9127
15	Positive	0.7197	0.6829	0.3120	0.9125
16	Positive	0.6613	0.6186	0.3153	0.9137
17	Positive	0.6673	0.6252	0.3150	0.9136
18	Positive	0.7853	0.7558	0.3084	0.9111
19	Positive	0.6724	0.6307	0.3147	0.9135
20	Positive	0.3000	0.2338	0.3355	0.9207
21	Negative	0.6135	0.5664	0.3180	0.9147
22	Negative	0.2064	0.1374	0.3408	0.9224
23	Positive	0.1943	0.1250	0.3414	0.9226
24	Positive	0.1961	0.1269	0.3413	0.9226
Test scale				0.3193	0.9184

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press

Table 3 Factor loading analysis for the objective 1

Factor	Eigenvalue	Difference	Proportion	Cumulative
Item 1	9.32370	7.49413	0.3885	0.3885
Item 2	1.82957	0.56318	0.0762	0.4647
Item 3	1.26639	0.02772	0.0528	0.5175
Item 4	1.23867	0.14194	0.0516	0.5691
Item 5	1.09673	0.21548	0.0457	0.6148
Item 6	0.88125	0.07713	0.0367	0.6515
Item 7	0.80412	0.03120	0.0335	0.6850
Item 8	0.77292	0.07751	0.0322	0.7172
Item 9	0.69541	0.04232	0.0290	0.7462
Item 10	0.65309	0.05160	0.0272	0.7734
Item 11	0.60150	0.06530	0.0251	0.7985
Item 12	0.53620	0.02074	0.0223	0.8208
Item 13	0.51546	0.03069	0.0215	0.8423
Item 14	0.48477	0.03638	0.0202	0.8625
Item 15	0.44839	0.03725	0.0187	0.8812
Item 16	0.41114	0.01019	0.0171	0.8983
Item 17	0.40094	0.01877	0.0167	0.9150
Item 18	0.38217	0.03639	0.0159	0.9309
Item 19	0.34578	0.02949	0.0144	0.9453
Item 20	0.31630	0.02441	0.0132	0.9585
Item 21	0.29189	0.02009	0.0122	0.9707
Item 22	0.27180	0.04083	0.0113	0.9820
Item 23	0.23097	0.03011	0.0096	0.9916
Item 24	0.20085	—	0.0084	1.0000

Likelihood ratio test statistic: independent *v.* saturated: χ^2 (276) = 3850.52, $P > \chi^2$ = 0.0000.

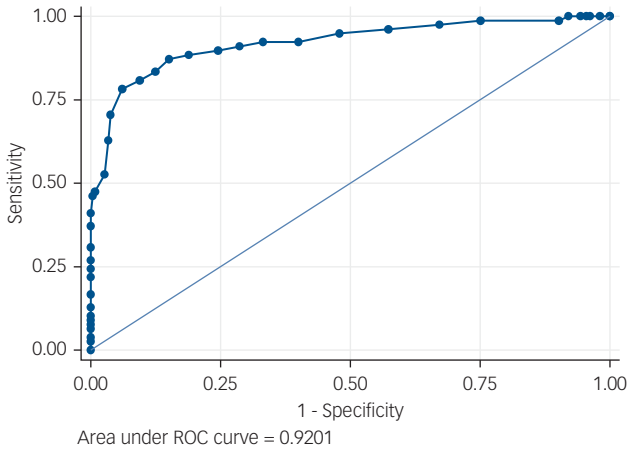

Fig. 1 Receiver operating characteristics of the AViDI-18. Observations = 342, AUC = 0.9201, s.e. = 0.0203, CI = 0.88–0.95. ROC, receiver operating characteristics.

The mean item score of the 24 items was 23.5 (s.d. ± 11.7) (Table 1). In the multivariable analysis, younger (β = 0.17, CI = 0.05–0.29, $P = 0.003$), unemployed (β = 0.65, CI = 0.13–1.17, $P = 0.013$) and poorly educated (β, −3.11, CI = −4.47 to −1.44, $P < 0.001$) participants were more likely to have higher scores (Table 1).

The average inter-item correlation of the 24 items was 0.32, with a Cronbach's α of 0.92 (Table 2). The inter-rater reliability agreement between two research assistants ($n = 30$) was 92.1% (expected agreement 77.9%, weighted kappa 0.64, $P < 0.001$). The test-retest reliability score was 0.86 (Spearman's correlation coefficient). Four factors with eigenvalues greater than 1.2 were retained at factor loading analysis (Table 3).

Of the 24 items, 6 were removed due to their ambiguity at face and content validity assessment (steps ii–iv) as well as due to poor item correlations and factor loading (step v). The items removed were items 7, 19, 20 and 21 (meant to depict happy, neutral or euthymic states) as well as item 8 (meant to depict a sad/worried person) and item 22 (was meant to depict a person who does not intend to die by suicide). These items are attached as a Supplementary Appendix, available at https://doi.org/10.1192/bjp.2018.9.

Table 4 Factor loading analysis of the final scale items (objective 2)

Factor	Eigenvalue	Difference	Proportion	Cumulative
Item 1	5.89906	4.67920	0.7227	0.7227
Item 2	1.21985	0.49610	0.1494	0.8721
Item 3	0.72376	0.13345	0.0887	0.9608
Item 4	0.59031	0.18888	0.0723	1.0331
Item 5	0.40143	0.15538	0.0492	1.0823
Item 6	0.24605	0.07420	0.0301	1.1124
Item 7	0.17185	0.05701	0.0211	1.1335
Item 8	0.11484	0.05839	0.0141	1.1476
Item 9	0.05645	0.03563	0.0069	1.1545
Item 10	0.02082	0.04679	0.0026	1.1570
Item 11	−0.02597	0.03673	−0.0032	1.1539
Item 12	−0.06270	0.06768	−0.0077	1.1462
Item 13	−0.13038	0.01234	−0.0160	1.1302
Item 14	−0.14272	0.05140	−0.0175	1.1127
Item 15	−0.19411	0.01026	−0.0238	1.0889
Item 16	−0.20438	0.03990	−0.0250	1.0639
Item 17	−0.24427	0.03300	−0.0299	1.0340
Item 18	−0.27728	0	−0.0340	1.0000

Objective 2: validating (assessing the sensitivity and specificity) the AViDI-18

A sample of 343 participants (97% of the target sample size of 353) in Uganda and South Africa was obtained. Of the 343 participants, 116 (33.8%) were male. Participant mean age was 35.32 years (s.d. ± 12.3). Over one-third (131, 38.2%) of the participants had low levels of education (no formal education or only up to primary level education). Just over one-quarter (80, 25.5%) of the participants were HIV-positive (Table 4).

Table 5 Cut-off scores of the AViDI-18

Cut-off point	Sensitivity (%)	Specificity (%)	Correctly classified (%)	LR+	LR−
≥0	100	0.00	22.74	1.00	0.0000
≥5	100.00	7.92	28.86	1.0861	0.0000
≥6	98.72	9.81	30.03	1.0946	0.1307
≥7	98.72	24.91	41.69	1.3146	0.0515
≥8	97.44	32.83	47.52	1.4506	0.0781
≥9	96.15	42.64	54.81	1.6764	0.0902
≥10	94.87	52.08	61.81	1.9796	0.0985
≥11	92.31	60.00	67.35	2.3077	0.1282
≥12	92.31	66.79	72.59	2.7797	0.1152
≥13	91.03	71.32	75.80	3.1739	0.1258
≥14	89.74	75.47	78.72	3.6588	0.1359
≥15	88.46	81.13	82.80	4.6885	0.1422
≥16	87.18	84.91	85.42	5.7756	0.1510
≥17	83.33	87.55	86.59	6.6919	0.1904
≥18	80.77	90.57	88.34	8.5615	0.2123
≥19	78.21	93.96	90.38	12.9527	0.2320
≥20	70.51	96.23	90.38	18.6859	0.3064
≥21	62.82	96.60	88.92	18.4972	0.3849
≥22	52.56	97.36	87.17	19.8993	0.4872
≥23	47.44	99.25	87.46	62.8527	0.5296
≥25	46.15	99.62	87.46	122.3080	0.5405
≥26	41.03	100.00	86.59		0.5897
≥44	0.00	100.00	77.26%		1.0000

Scores ranging from 0 to 5 have been omitted from the table because the scale is 100% sensitive, and those ranging from 27 to 44 are omitted because the scale is 100% specific. LR+ positive likelihood ratio. LR− Negative Likelihood ratio. A score of ≥16 is the scale's best cut-off point.

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press

Table 6 Scale accuracy by sociodemographic variables

Study variable	Reference	ROC area	Level of significance
Gender	Male (116) Female (227)	0.89 (0.78–0.99) 0.93 (0.89–0.97)	$\chi^2 = 0.59$, $P = 0.440$
Age	18–30 (146) 31–49 (144) 50–84 (53)	0.90 (0.83–0.97) 0.94 (0.88–1.00) 0.89 (0.80–0.99)	$\chi^2 = 0.96$, $P = 0.618$
HIV sero-status	HIV-positive (80) HIV-negative (263)	0.94 (0.89–1.00) 0.93 (0.89–0.97)	$\chi^2 = 0.15$, $P = 0.698$
Education status	Up to primary level education (131) At least secondary level education (212)	0.93 (0.89–0.98) 0.90 (0.83–0.96)	$\chi^2 = 0.76$, $P = 0.384$
Employment	Formal employment (112) Non-formal employment (82) Unemployed (149)	0.94 (0.88–1.00) 0.86 (0.77–0.96) 0.94 (0.89–0.98)	$\chi^2 = 2.07$, $P = 0.354$

ROC, receiver operating characteristic.

The mean item score of the AViDI-18 was 13.23 (s.d. ± 8.2), with an average inter-item covariance of 0.28 and a Cronbach's α of 0.87. The mean time for the scale administration was 10 min (s.d. ± 5.2).

The prevalence of a DSM-based MDD episode according to the MINI was 22.7% (78/343). The AUC score of the AViDI-18 was 0.9 (0.88–0.95), with an s.e. of 0.02 (Fig. 1). A cut-off score of 16 produced the best balance between sensitivity, specificity and likelihood ratios with a sensitivity of 87%, specificity of 84.9%, a negative likelihood ratio of 5.8 and a positive likelihood ratio of 0.155. At this cut-off score, the scale was able to correctly identify 85.4% of cases. The negative predictive value (NPV) of the scale was 77.7% (CI 72.9–82%), with a positive predictive value (PPV) of 100% (CI 15.8–100%) (Table 5).

There were no statistically significant differences in scale accuracy by gender ($\chi^2 = 0.15$, $P = 0.59$), age categories ($\chi^2 = 0.96$, $P = 0.62$), level of education ($\chi^2 = 0.76$, $P = 0.38$), employment status ($\chi^2 = 2.07$, $P = 0.35$) and HIV sero-status ($\chi^2 = 0.15$, $P = 0.70$) (Table 6).

Receiver operating characteristics were similar in South Africa and in Uganda. In South Africa, the AUC score for the AViDI-18 was 0.92 (0.86–0.98), with a PPV of 100% (2.5–100%) and a NPV of 89.6% (83.2–94.2%) (Fig. 2). In Uganda, the AUC score was 0.92 (0.86–0.96), with a PPV of 100% (2.5–100%) and a NPV of 69.9% (63.14–76.08%) (Fig. 3).

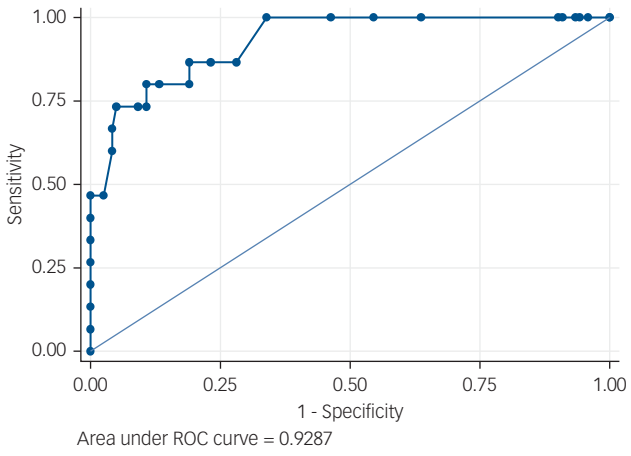

Fig. 2 Receiver operating characteristics curve for South Africa. Observations = 136, AUC = 0.92, s.e. 0.03, CI = 0.86–0.98. ROC, receiver operating characteristics.

Discussion

Our study is among the first to demonstrate that a visual scale can be used as an accurate screener for a DSM-based MDD in patients with rates of low literacy.[5,6]

The high accuracy of the AViDI-18 compared with previous visual scales is notable,[5,6] and may reflect a number of factors. First, we used multiple signs and symptoms including mood, appetite, sleep, energy, worry and suicidality to assess for depression; whereas a number of previous visual scales have used single items of sadness as a measure of depression,[7–10] so limiting their sensitivity and specificity.[11,12] Second, we conducted extensive face and content validity studies of the items followed by a factor loading analysis before including them into the final scale, whereas the development of previous visual scales did not incorporate such steps.[7–9] Third, our scale consisted of coloured line drawings, whereas a number of previous visual scales are comprised of black and white shaded drawings. Coloured pictures are easier to understand and more appealing, especially in persons with low literacy.[21] Fourth, the addition of simple lettered instructions to pictures improve understanding compared with text alone, and so may have enhanced the ability of partially literate participants to fully appreciate the items and provide appropriate responses.[22]

We found that scale accuracy was not significantly associated with sociodemographic variables or HIV status. Previous work has suggested that the meanings individuals attach to pictures may differ by gender.[23,24] The findings here suggest that pictorial items may, however, be an effective means of communication for a broad range of individuals with low literacy.[22,25] It is also notable that there was no significant difference in the accuracy of the AViDI-18 by HIV status, even though physical symptoms of AIDS may overlap with the signs and symptoms of depression. This finding confirms the potential value of the AViDI-18 as a screening tool for depression, even in populations with comorbid physical illness.

However, a number of limitations do deserve emphasis. First, AViDI-18 items were based on the DSM and the scale was validated against the MINI, rather than compared with a longitudinal expert assessment and diagnosis gold standard. Moreover, it was not possible to depict some of the DSM-based criteria for depression including guilt, psychomotor retardation or concentration pictorially without creating ambiguity. Second, our findings indicate that although the PPV of the AViDI-18 was very high, the confidence interval for this value was very wide too. The wide confidence interval of the PPV means that caution needs to be taken when interpreting a positive screen result. Since a high prevalence of a disease predicts an equally high PPV, and the prevalence of MDD in our

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press

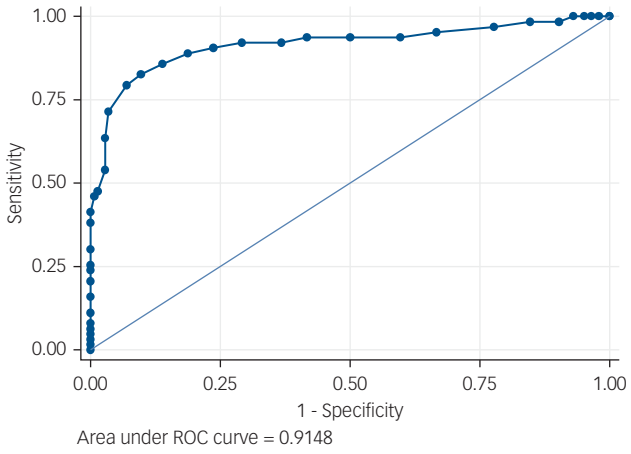

Fig. 3 Receiver operating characteristics curve for Uganda. Observations = 207, AUC = 0.91, s.e. = 0.02, CI = 0.86–0.96.

Note: Please contact the authors for information regarding the actual instructions for scale administration, scoring of the items and the exact picture scales (measurements) to use.

study sample was high (22%), the AViDI-18 requires validation in a population with a low depression prevalence. Furthermore, given concerns about the value of screening for depression, it is also important to ascertain whether screening with the AViDI-18 leads to improved diagnosis and ultimately better patient outcomes. The sample sizes for future studies will need to be calculated for both specificity and sensitivity and is likely to be larger. Lastly, at factor loading, two factors with eigenvalues greater than 1.2 were identified. This finding indicates that the AViDI-18 could be measuring constructs other than depression. However, factor 1 had an eigenvalue of 5.9, so we can be confident that the AViDI-18 measured DSM-based MDD, perhaps with other potential comorbidities. The study sample size of 343 (97% of the target sample size) fell short by 10 participants, which could have reduced the power to detect statistically significant differences. Our findings should be interpreted cautiously with this in mind.

In conclusion, the findings from our study demonstrated that visual scales can be used to accurately detect depression in patients with low literacy in low-resourced countries in sub-Saharan Africa. It may be useful for further work to compare the accuracy of the AViDI-18 against established screening instruments such as the PHQ-9. We are also aware of a number of novel methods of administering screening instruments to individuals with low literacy, such as mobile phones and tablet computers with visual cues and auditory instructions, and we recommend that future studies explore these methods.

Dickens Akena, MBChB, M.Med, PhD, Department of Psychiatry, Makerere University College of Health Sciences, Kampala, Uganda and Department of Psychiatry and Mental Health, Faculty of Health Sciences, University of Cape Town, South Africa; **John Joska**, MBChB, M.Med, PhD, **Dan J. Stein**, MBChB, FRCP, PhD, Department of Psychiatry and Mental Health, Faculty of Health Sciences, University of Cape Town, South Africa.

Correspondence: Dickens Akena, MBChB, M.Med, PhD, P.O.Box 16456 Wandegya, Kampala, Uganda. Email: akenadickens@yahoo.co.uk

First received 1 Jun 2017, final revision 14 Dec 2017, accepted 24 Dec 2017

Supplementary material

Supplementary material is available online at https://doi.org/10.1192/bjp.2018.9.

Funding

This work was supported by the African Research Excellence Fellowship fund (AREF-MF-001; principal investigator: D.A.). During the writing of the paper, D.A. received protected time from a capacity building NURTURE fellowship (grant number D43TW010132) supported by Office of the Director, National Institutes of Health; National Institute of Dental and Craniofacial Research; National Institute of Neurological Disorders and Stroke; National Heart, Lung and Blood Institute; Fogarty International Center and the National Institute on Minority Health and Health Disparities.

Ethical approvals

This study was approved by the Makerere School of Medicine Research Ethical Committee (reference 2015-113), the Uganda National Council for Science and Technology (reference SS 3963) and the Faculty of Health Sciences Human Research Ethics Committee at the University of Cape Town (reference 228/2016).

Author contributions

D.A. conceptualised this work and supervised the data collection and analysis. J.J. and D.S. provided extensive comments and critiques of the work at all the stages, from writing the proposal to the final manuscript. All three authors actively participated in writing the manuscript

References

1 Hamilton M. A rating scale for depression. *Jf Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press

2 Radloff LS. The CES-D scale: a self report depression scale for research in the general population. *Appl Psychol Meas* 1977; **1**: 385–401.

3 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9 validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–13.

4 UNESCO Institute for Statistics. Adult and Youth Literacy: National, Regional and Global trends, 1985–2015. Montreal, Quebec, Canada: UNESCO, 2013.

5 Akena D, Joska J, Musisi S, Stein DJ. Sensitivity and specificity of a visual depression screening instrument among HIV-positive individuals in Uganda, an area with low literacy. *AIDS Behav* 2012; **16**: 2399–406.

6 Puertas G, Patel V, Marshall T. Are visual measures of mood superior to questionnaire measures in non–Western settings? *Soc Psychiatry Psychiatr Epidemiol* 2004; **39**: 662–6.

7 Berg A, Lönnqvist J, Palomäki H, Kaste M. Assessment of depression after stroke a comparison of different screening instruments. *Stroke* 2008; **40**: 523–9.

8 Aitken RCB. Measurement of feelings using visual analogue scales. *Proc Roy Soc Med* 1969; **62**: 989–93.

9 Kertzman S, Aladjem Z, Milo R, Ben-Nahum Z, Birger M, Grinspan H, et al. The utility of the visual analogue scale for the assessment of depressive mood in cognitively impaired patients. *Int J Geriatr Psychiatry* 2004; **19**: 789–96.

10 Price CI, Curless RH, Rodgers H. Can stroke patients use visual analogue scales? *Stroke* 1999; **30**: 1357–61.

11 Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med* 2010; **8**: 348–53.

12 Maurer DM. Screening for depression. *Am Fam Physician* 2012; **82**: 139–44.

13 Sheehan DV, Lecrubier Y, Harnett-Sheehan K. The mini international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview. *J Clin Psychiatry* 1998; **59**: 22–33.

14 Marie-Laurence Flahaux and Hein De Haas. African migration: trends, patterns, drivers. *Comparative Migration Studies* 2016; **4** (1): https://doi.org/10.1186/s40878-40015-40015-40876.

15 Charter RA. Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *J Clin Expe Neuropsychol* 1999; **21**: 559–66.

16 Hobart JC, Cano SJ, Warner TT, Thompson AJ. What sample sizes for reliability and validity studies in neurology? *J Neurol* 2012; **259**: 2681–94.

17 Fenn Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996; **3**: 895–900.

18 Nakimuli-Mpungu E, Bass JK, Alexandre P, Mills EJ, Musisi S, Ram M, et al. Depression, alcohol use and adherence to antiretroviral therapy in sub-Saharan Africa: a systematic review. *AIDS Behav* 2011; **15**: 376–88.

19 Lawshe CH. A quantitative approach to content validity. *Personnel Psychol* 1975; **28**: 563–75.

20 Rahn M. Factor Analysis: A Short Introduction, Part 1. The Analysis Factor, 2016. Available at http://wwwtheanalysisfactorcom/factor-analysis-1-introduction/.

21 Readence JE, Moore DW. A meta-analytic review of the effect of adjunct pictures on reading comprehension. *Psychol Sch* 1981; **18**: 218–24.

22 Ngoh LN, Shepherd MD. Design, development, and evaluation of visual aids for communicating prescription drug instructions to nonliterate patients in rural Cameroon. *Patient Educ Couns* 1997; **3**: 245–61.

23 Salkind L, Salkind NJ. Gender and age differences in preference for works of art. *Studies in Art Education* 1997; **38**: 246–56.

24 Cela-Conde CJ, Ayala FJ, Munar E, Maestú F, Nadal M, Capó MA, et al. Sex-related similarities and differences in the neural correlates of beauty. *PNAS* 2009; **106**: 3847–52.

25 Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006; **61**: 173–90.

https://doi.org/10.1192/bjp.2018.9 Published online by Cambridge University Press